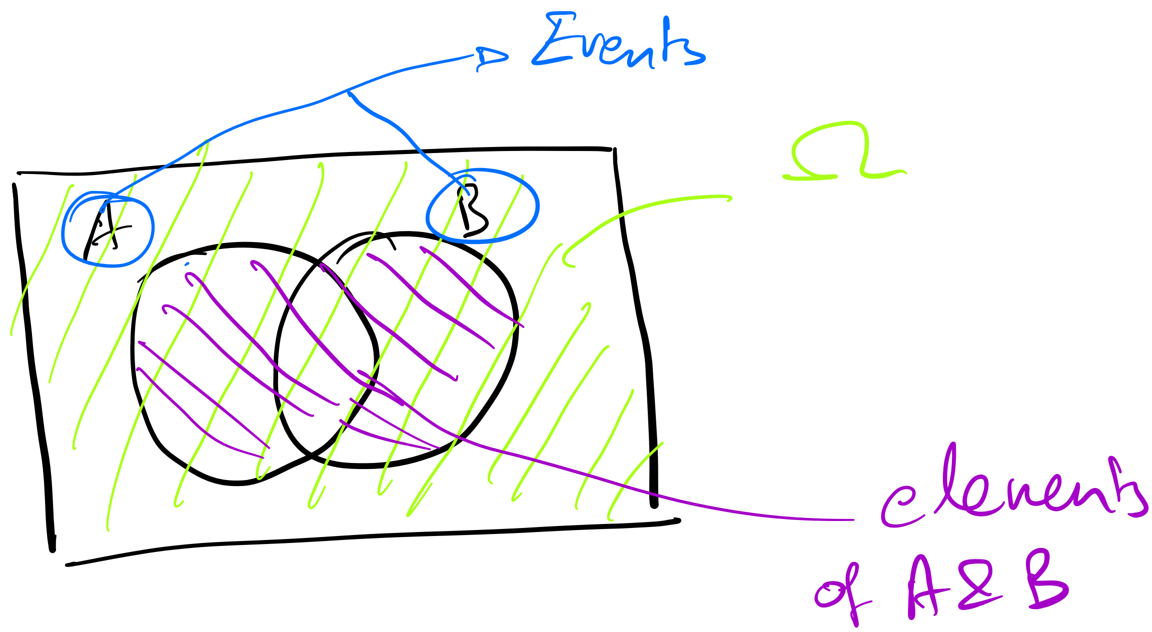
Chapter 1 - Probability

1.2 - Sample Spaces and Events

- Ω : sample space → set of possible outcomes of an experiment.
- ω : sample outcomes/realisations/elements

↪ Subset of Ω are called Events

eg



eg

Given an event A,

let $A^c = \{\omega \in \Omega : \omega \notin A\}$

↳ denotes the complement of A [A^c]

- The complement of Ω [sample space] is the empty set ϕ

Note:

- $A \cup B = \{\omega \in \Omega : \omega \in A$ or $\omega \in B$ or both$\}$

[A or B] [UNION]

if A_1, A_2, A_3 is a sequence of sets then:

$$\bigcup_{i=1}^{\infty} A_i = \{\omega \in \Omega : \omega \in A_i \text{ for at least one } i\}$$

- $A \cap B = \{\omega \in \Omega : \omega \in A$ and $\omega \in B\}$

[A and B] [INTERSECTION]

sometimes read as $\boxed{AB \text{ or } (A,B)}$

If $A_1, A_2 \ldots$ is a sequence of sets then

$$\bigcap_{i=1}^{\infty} A_i = \{\omega \in \Omega : \omega \in A_i \text{ for all } i\}$$

Note:

- The set difference is defined by
  $A - B = \{\omega : \omega \in A, (\omega \notin B)\}$ (not)
- If every element of A is contained in B
  $\rightarrow A \subset B$ or $B \supset A$
- If A is a finite set, let $|A|$ denote the number elements in A.

Summary of Terminology

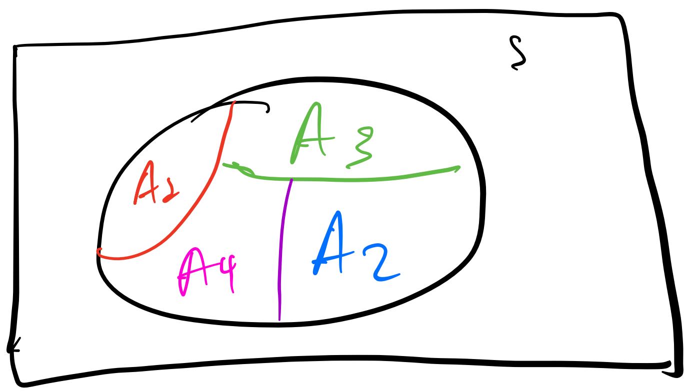
Ω	sample space
ω	outcome (point or element)
A	event (subset of Ω)
A^c	complement of A (not A)
$A \bigcup B$	union (A or B)
$A \bigcap B$ or AB	intersection (A and B)
$A - B$	set difference (ω in A but not in B)
$A \subset B$	set inclusion
$\emptyset$	null event (always false)
Ω	true event (always true)

- Disjoint or mutually exclusive

$[A \cap B = \emptyset]$ $A \neq B$

- A partition of Ω is a sequence of disjointed sets A_1, A_2 such that

$\bigcup_{i=1}^{\infty} A_i = \Omega$: eg $A_1, A_2, A_3 \ldots$

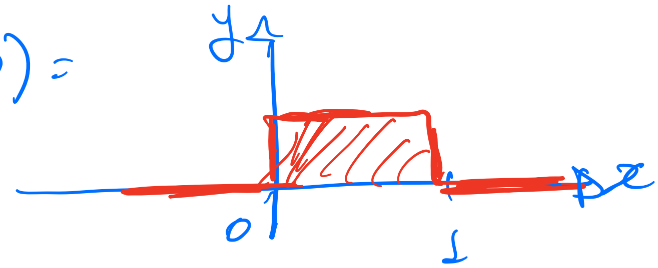


* Given an event A, define the indicator function of A by:

$$\mathbb{I}_A(\omega) = I(\omega \in A) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

eg $\mathbb{I}_{[0,1]}(\omega) =$

$(y = x)$



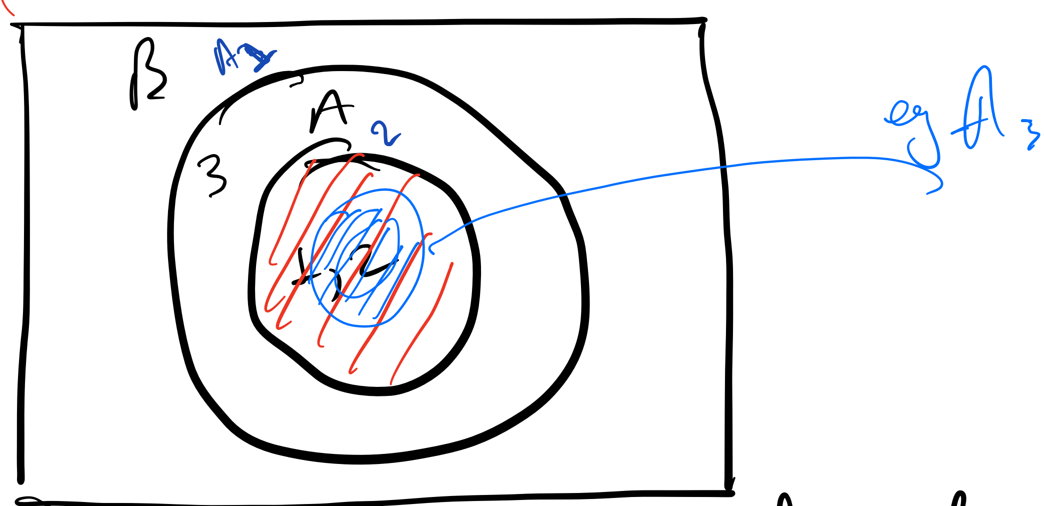
- A sequence of sets $A_1, A_2, \dots,$ is monotone increasing if $A_1 \subset A_2 \subset A_3 \dots$
and we define $\lim_{n\to\infty} A_n = \bigcup_{i=1}^{\infty} A_i$

A sequence of sets $A_1, A_2, \dots,$ is monotone decreasing if $A_1 \supset A_2 \supset A_3$
and we define $\lim_{n\to\infty} A = \bigcap_{i=1}^{\infty} A_i$

$A_n \to A$

4

(monotone decreasing)



Elements which is the intersection of $\underline{A \cap B}$ are $\{1,2\}$, because 3 is <u>not</u> included into B

1.3 - Probability

- $\mathbb{P}(A) \rightsquigarrow$ probability of A
  - ↳ $\mathbb{P} \rightsquigarrow$ probability distribution or probability measure

* To qualify as a probability, $\mathbb{P}$ must satisfy 3 AXIOMS

1.5 Definition. *A function $\mathbb{P}$ that assigns a real number $\mathbb{P}(A)$ to each event A is a* **probability distribution** *or a* **probability measure** *if it satisfies the following three axioms:*

Axiom 1: $\mathbb{P}(A) \geq 0$ *for every* A

Axiom 2: $\mathbb{P}(\Omega) = 1$

Axiom 3: *If* $A_1, A_2, \ldots$ *are disjoint then*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

[1] It is not always possible to assign a probability to every event A if the sample space is large, such as the whole real line. Instead, we assign probabilities to a limited class of set called a σ-field. See the appendix for details.

↳ What does this exactly mean??
{Need to do some more research}

1.9 Appendix

Generally, it is not feasible to assign probabilities to all subsets of a sample space Ω. Instead, one restricts attention to a set of events called a **σ-algebra** or a **σ-field** which is a class $\mathcal{A}$ that satisfies:

(i) $\emptyset \in \mathcal{A}$,

(ii) if $A_1, A_2, \ldots, \in \mathcal{A}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ and

(iii) $A \in \mathcal{A}$ implies that $A^c \in \mathcal{A}$.

The sets in $\mathcal{A}$ are said to be **measurable**. We call $(\Omega, \mathcal{A})$ a **measurable space.** If $\mathbb{P}$ is a probability measure defined on $\mathcal{A}$, then $(\Omega, \mathcal{A}, \mathbb{P})$ is called a **probability space.** When Ω is the real line, we take $\mathcal{A}$ to be the smallest σ-field that contains all the open subsets, which is called the **Borel σ-field**.

* Interpretations of $\mathbb{P}(A)$

- Frequency: $\mathbb{P}(A)$ is the long run proportion (tend to) of times that A is true in repetition
[eg: Flipping a coin many times $\mathbb{P}(H) \to 1/2$]
- Degree of belief: $\mathbb{P}(A)$ measures an observer's strength of belief that A is true

Both cases AXIOM ① & ③ HOLD

① $\mathbb{P}(A) \geq 0$

③ If disjoint $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

One can derive many properties of $\mathbb{P}$ from the axioms, such as:

$$\begin{aligned}
\mathbb{P}(\emptyset) &= 0 \\
A \subset B &\implies \mathbb{P}(A) \leq \mathbb{P}(B) \\
0 \leq &\ \mathbb{P}(A) \leq 1 \\
\mathbb{P}(A^c) &= 1 - \mathbb{P}(A) \\
A \bigcap B = \emptyset &\implies \mathbb{P}\left(A \bigcup B\right) = \mathbb{P}(A) + \mathbb{P}(B).
\end{aligned} \tag{1.1}$$

A less obvious property is given in the following Lemma.

1.6 Lemma. *For any events A and B,*

$$\mathbb{P}\left(A \bigcup B\right) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB).$$

PROOF. Write $A \bigcup B = (AB^c) \bigcup (AB) \bigcup (A^c B)$ and note that these events are disjoint. Hence, making repeated use of the fact that $\mathbb{P}$ is additive for disjoint events, we see that

$$\begin{aligned}
\mathbb{P}\left(A \bigcup B\right) &= \mathbb{P}\left((AB^c) \bigcup (AB) \bigcup (A^c B)\right) \\
&= \mathbb{P}(AB^c) + \mathbb{P}(AB) + \mathbb{P}(A^c B) \\
&= \mathbb{P}(AB^c) + \mathbb{P}(AB) + \mathbb{P}(A^c B) + \mathbb{P}(AB) - \mathbb{P}(AB) \\
&= P\left((AB^c) \bigcup (AB)\right) + \mathbb{P}\left((A^c B) \bigcup (AB)\right) - \mathbb{P}(AB) \\
&= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB). \quad \blacksquare
\end{aligned}$$

eg 2 coin tosses.

$H_1 = \mathbb{P}(H)$ - 1st toss

$H_2 = \mathbb{P}(H)$ - 2nd toss

$\therefore \mathbb{P}(H_1 \cup H_2) = \mathbb{P}(H_1) + \mathbb{P}(H_2) - \mathbb{P}(H_1 H_2)$

$= (1/2) + (1/2) - (1/2 \times 1/2) = \underline{\underline{3/4}}$

1.8 Theorem (Continuity of Probabilities). *If $A_n \to A$ then*

$$\mathbb{P}(A_n) \to \mathbb{P}(A)$$

as $n \to \infty$.

Proof. Suppose that A_n is monotone increasing so that $A_1 \subset A_2 \subset \cdots$. Let $A = \lim_{n\to\infty} A_n = \bigcup_{i=1}^{\infty} A_i$. Define $B_1 = A_1$, $B_2 = \{\omega \in \Omega : \ \omega \in A_2, \omega \notin A_1\}$, $B_3 = \{\omega \in \Omega : \ \omega \in A_3, \omega \notin A_2, \omega \notin A_1\}, \ldots$ It can be shown that $B_1, B_2, \ldots$ are disjoint, $A_n = \bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} B_i$ for each n and $\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$. (See exercise 1.) From Axiom 3,

$$\boxed{\mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{i=1}^{n} B_i\right) = \sum_{i=1}^{n} \mathbb{P}(B_i)}$$

and hence, using Axiom 3 again,

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(B_i) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \mathbb{P}(A). \quad \blacksquare$$

↳ Check out pg 5 of Notes

1.4 - Probability on Finite Sample Spaces

If Ω is finite and if each outcome is equally likely, then

$$\boxed{\mathbb{P}(A) = \frac{|A|}{|\Omega|},}$$

* This is called uniform probability distribution.

eg Suppose we toss a die twice.
Then $\Omega = \{(i,j) ; i,j \in \{1, \dots, 6\}\}$
[has 36 elements $(6 * 6)$]

- If each outcome is equally likely,
then $\mathbb{P}(A) = \frac{|A|}{36}$ → $|A|$ denotes the number of elements in A.
→ 36 : Ω (sample space)

- The probability that the sum of the dice = 11 means [5, 6] [6, 5] } 2
→ is : 2/36

* Methods for counting points are called combinatorial methods

*Counting Theory (FACT)

- Given n objects
- The number of ways of ordering these objects is

$$n! = n(n-1)(n-2) \cdots 3 \times 2 \times 1$$

$\boxed{\text{*Note: } 0! = 1}$

↳ $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$ } "n choose k"

eg If we have a class of 20 people, and we select a committee of 3 students, then there are

$$\binom{20}{3} = \frac{20!}{3!(20-3)!} = \frac{20 \times 19 \times 18}{3 \times 2 \times 1} = 1140$$

∴ There are 1140 possible committees.

* Propriétés :

- $\binom{n}{0} = \binom{n}{n} = 1$

- $\binom{n}{k} = \binom{n}{n-k}$

1.5 - Independent Events

1.9 Definition. *Two events A and B are* **independent** *if*

$$\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B) \tag{1.3}$$

and we write $A \amalg B$. A set of events $\{A_i : i \in I\}$ is independent if

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i)$$

for every finite subset J of I. If A and B are not independent, we write

$$A \rightsquigarrow B$$

eg If we flip a fair coin twice, then the probability of 2 heads is $1/2 \times 1/2$

* Independence can either be assumed or derive → verifying that $P(AB) = P(A)P(B)$ holds

↳ Tossing a coin twice

eg Tossing a fair die

Let $A = \{2, 4, 6\}$

Let $B = \{1, 2, 3, 4\}$

$A \cap B = \{2, 4\}$

$\therefore \mathbb{P}(AB) = 2/6 = \mathbb{P}(A)\,\mathbb{P}(B) = \left(\frac{1}{2}\right) \times \left(2/3\right)$

$\hookrightarrow \therefore$ events are independent

Note:

$\&$ Suppose A & B are disjoint events, each with positive probability.

$\to$ Can they be independent? No

$\hookrightarrow \mathbb{P}(A)\,\mathbb{P}(B) > 0$ and $\mathbb{P}(A \cap B) = \mathbb{P}(\phi) = 0$
$(\mathbb{P}(AB))$

[special case]

Summary of Independence

1. A and B are independent if and only if $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$.
2. Independence is sometimes assumed and sometimes derived.
3. Disjoint events with positive probability are not independent.

Examples:

① Tossing a fair coin 10 times.

Let A = "at least 1 Head"

Let T_j = tails occur on the j^{th} toss

$\therefore$

$$\begin{aligned}\mathbb{P}(A) &= 1 - \mathbb{P}(A^c) \\ &= 1 - \mathbb{P}(\text{tails}) \\ &= 1 - \mathbb{P}(T_1 T_2 T_3 \ldots T_{10}) \\ &= 1 - \mathbb{P}(T_1)\,\mathbb{P}(T_2) \ldots \mathbb{P}(T_{10}) \\ &= 1 - (1/2)^{10} \simeq 0.999\end{aligned}$$

2 * Important *

1.11 Example. Two people take turns trying to sink a basketball into a net. Person 1 succeeds with probability 1/3 while person 2 succeeds with probability 1/4. What is the probability that person 1 succeeds before person 2? Let E denote the event of interest. Let A_j be the event that the first success is by person 1 and that it occurs on trial number j. Note that $A_1, A_2, \ldots$ are disjoint and that $E = \bigcup_{j=1}^{\infty} A_j$. Hence,

$$\mathbb{P}(E) = \sum_{j=1}^{\infty} \mathbb{P}(A_j).$$

Now, $\mathbb{P}(A_1) = 1/3$. A_2 occurs if we have the sequence person 1 misses, person 2 misses, person 1 succeeds. This has probability $\mathbb{P}(A_2) = (2/3)(3/4)(1/3) = (1/2)(1/3)$. Following this logic we see that $\mathbb{P}(A_j) = (1/2)^{j-1}(1/3)$. Hence,

$$\boxed{\mathbb{P}(E) = \sum_{j=1}^{\infty} \frac{1}{3}\left(\frac{1}{2}\right)^{j-1} = \frac{1}{3}\sum_{j=1}^{\infty}\left(\frac{1}{2}\right)^{j-1} = \frac{2}{3}.}$$

Here we used that fact that, if $0 < r < 1$ then $\sum_{j=k}^{\infty} r^j = r^k/(1-r)$. ■

eg $M_1 M_2 S_1 \rightarrow (1/2)(1/3)$

$A_3 \rightarrow M_1 M_2 M_1 M_2 S_1 \quad (2/3)(3/4)(2/3)(3/4)$

$\rightarrow (j-1) \rightarrow (\frac{1}{2})(\frac{1}{2})(1/3)$

$\therefore (1/3)$ is always present

$\therefore (1/2)^{j-1}$

1.6 - Conditional Probability

Assuming that $\mathbb{P}(B) > 0$, we define the conditional probability of A given that B has occurred as follows:

1.12 Definition. *If $\mathbb{P}(B) > 0$ then the* **conditional probability** *of A given B is*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}. \tag{1.4}$$

* Note: Think $\mathbb{P}(A|B)$ as the fraction of times A occurs among those in which B occurs

- $\mathbb{P}(\cdot|B)$ = is a probability that satisfies all 3 Axioms
- $\mathbb{P}(\Omega|B) = 1$
- If disjoint: $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i|B\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i|B)$

$(A_1, A_2, A_3 \ldots)$

BUT - in general NOT true

- $\mathbb{P}(A|B \cup C) = \mathbb{P}(A|B) + \mathbb{P}(A|C)$
- $\boxed{\mathbb{P}(A|B) \neq \mathbb{P}(B|A)}$

eg The probability of spots given you have measles is 1

BUT: probability that you have measles given that you have spots $\neq 1$

* Example 1.13 (Important) *

1.13 Example. A medical test for a disease D has outcomes $+$ and $-$. The probabilities are:

	D	D^c
$+$	.009	.099
$-$	.001	.891

From the definition of conditional probability,

$$\mathbb{P}(+|D) = \frac{\mathbb{P}(+\bigcap D)}{\mathbb{P}(D)} = \frac{.009}{.009 + .001} = .9$$

and

$$\mathbb{P}(-|D^c) = \frac{\mathbb{P}(-\bigcap D^c)}{\mathbb{P}(D^c)} = \frac{.891}{.891 + .099} \approx .9.$$

Apparently, the test is fairly accurate. Sick people yield a positive 90 percent of the time and healthy people yield a negative about 90 percent of the time. Suppose you go for a test and get a positive. What is the probability you have the disease? Most people answer .90. The correct answer is

$$\mathbb{P}(D|+) = \frac{\mathbb{P}(+\bigcap D)}{\mathbb{P}(+)} = \frac{.009}{.009 + .099} \approx .08.$$

The lesson here is that you need to compute the answer numerically. Don't trust your intuition. ■

The results in the next lemma follow directly from the definition of conditional probability.

1.14 Lemma. *If A and B are independent events then $\mathbb{P}(A|B) = \mathbb{P}(A)$. Also, for any pair of events A and B,*

$$\mathbb{P}(AB) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A).$$

From the last lemma, we see that another interpretation of independence is that knowing B doesn't change the probability of A. The formula $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B|A)$ is sometimes helpful for calculating probabilities.

Summary of Conditional Probability

1. If $\mathbb{P}(B) > 0$, then
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}.$$

2. $\mathbb{P}(\cdot|B)$ satisfies the axioms of probability, for fixed B. In general, $\mathbb{P}(A|\cdot)$ does not satisfy the axioms of probability, for fixed A.

3. In general, $\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$.

4. A and B are independent if and only if $\mathbb{P}(A|B) = \mathbb{P}(A)$.

1.7 - Baye's Theorem

→ Baye's theorem is the basics of "expert systems" & "Bayes' net",

1.7 Bayes' Theorem

Bayes' theorem is the basis of "expert systems" and "Bayes' nets," which are discussed in Chapter 17. First, we need a preliminary result.

1.16 Theorem (The Law of Total Probability). *Let $A_1, \ldots, A_k$ be a partition of Ω. Then, for any event B,*

$$\mathbb{P}(B) = \sum_{i=1}^{k} \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

PROOF. Define $C_j = BA_j$ and note that $C_1, \ldots, C_k$ are disjoint and that $B = \bigcup_{j=1}^{k} C_j$. Hence,

$$\mathbb{P}(B) = \sum_j \mathbb{P}(C_j) = \sum_j \mathbb{P}(BA_j) = \sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)$$

since $\mathbb{P}(BA_j) = \mathbb{P}(B|A_j)\mathbb{P}(A_j)$ from the definition of conditional probability. ■

~ rearange equation.

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(BA)}{\mathbb{P}(A)}$$

1.17 Theorem (Bayes' Theorem). *Let $A_1, \ldots, A_k$ be a partition of Ω such that $\mathbb{P}(A_i) > 0$ for each i. If $\mathbb{P}(B) > 0$ then, for each $i = 1, \ldots, k$,*

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)}. \tag{1.5}$$

1.18 Remark. We call $\mathbb{P}(A_i)$ the **prior probability of** A and $\mathbb{P}(A_i|B)$ the **posterior probability of** A.

PROOF. We apply the definition of conditional probability twice, followed by the law of total probability:

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_iB)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)}. \quad ■$$

1.19 Example. I divide my email into three categories: A_1 = "spam," A_2 = "low priority" and A_3 = "high priority." From previous experience I find that

$\mathbb{P}(A_1) = .7$, $\mathbb{P}(A_2) = .2$ and $\mathbb{P}(A_3) = .1$. Of course, $.7 + .2 + .1 = 1$. Let B be the event that the email contains the word "free." From previous experience, $\mathbb{P}(B|A_1) = .9$, $\mathbb{P}(B|A_2) = .01$, $\mathbb{P}(B|A_1) = .01$. (Note: $.9 + .01 + .01 \neq 1$.) I receive an email with the word "free." What is the probability that it is spam? Bayes' theorem yields,

$$\mathbb{P}(A_1|B) = \frac{.9 \times .7}{(.9 \times .7) + (.01 \times .2) + (.01 \times .1)} = .995. \quad ■$$